International Journal of Mechanical Engineering and Technology (IJMET)
Volume 10, Issue 01, January 2019, pp. 287–295, Article ID: IJMET_10_01_029
Available online at http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=1
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

A MODEL AND ALGORITHM FOR DETECTING SPYWARE IN MEDICAL INFORMATION SYSTEMS

V. Lakhno
Department of Computer Systems and Networks,
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine

D. Kasatkin
Associate Professor of the Department of Computer Systems and Networks,
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine

V. Kozlovskyi
Vice Rector of National Aviation University,
National Aviation University, Kyiv, Ukraine

S. Petrovska
Dean of the Faculty of Economics and Business Administration,
National Aviation University, Kyiv, Ukraine

Y. Boiko
Associate Professor of the Department of Applied Information Systems,
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

P. Kravchuk
Associate Professor of the Department of Security Management of an Enterprise,
Private Higher Educational Establishment “European University”, Kyiv, Ukraine

N. Lishchynovska
Graduate student, State University of Telecommunications, Kyiv, Ukraine

ABSTRACT

The article outlines extensions to the model and algorithm of spyware detection procedures which, in particular, presents a potential threat to medical information systems. The approaches and solutions in this paper allow to programmatically implement binary file segmentation using discrete wavelet transform (WT). Unlike the existing approaches, the model and algorithm proposed in the article took into account local extrema of the wavelet coefficients (WC). The described solutions allow for analysis of potentially dangerous and spyware files the number of bytes, as well as the entropy for individual segments of files that are transmitted for analysis to the

threat detection module. Additionally considered an alternative solution, in which the model was based on the procedure for analyzing the frequency of occurrence of file bytes in segments or the location of bytes in segments of the analyzed files. The wavelet transform was chosen based on maximizing efficiency for the analysis of spyware, that are potentially dangerous due to collecting confidential data in MIS. Experimental testing of the models was completed during the analysis of the keylogger file, as well as other spyware programs installed in medical information systems (MIS), that were not recognized as malicious by regular antiviruses. At the same time, during the experimental verification, the task was to minimize the requirements for computational resources on computers on which the MIS were running.

Key words: Cybersecurity, medical information system, spyware detection, wavelet transform, file segmentation.

Cite this Article: V. Lakhno, D. Kasatkin, V. Kozlovskyi, S. Petrovska, Y. Boiko, P. Kravchuk, N. Lishchynovska, A Model and Algorithm for Detecting Spyware in Medical Information Systems, *International Journal of Mechanical Engineering and Technology* 10(1), 2019, pp. 287–295.
http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=10&IType=1

1. INTRODUCTION

With the increased use of information technologies and systems (ITS) in medicine, technical specialists are facing the problem of ensuring their information security (IS) and data protection (DP) [1]. In practice, in an environment where clinics and other health facilities faced an increase in the number of personal data of patients and staff, and, especially for medical institutions, whose data is protected by medical confidentiality, the task of reliable cyber defense of medical information systems (MIS) against all sorts of destructive interference in their work by computer intruders or dishonest staff.

With the growing number and complexity of methods of destructive interference in the work of information systems around the world [2, 3], the task of ensuring IS and DP in MIS moved to a new level of quality. Despite the sufficient elaboration of the general methodology for the construction of MIS protection systems, local tasks still remain, for example, protection against the intrusion of “viruses” into software that record the patient’s medical history or store data from examinations and analyzes. Therefore, in our opinion, the task of developing models and algorithms that help detect spyware that has been pre-packaged (VPPU) remains an urgent. This problem, for example can be solved by analyzing, for all files, their entropy characteristics of the parameters. This approach ensures the minimization of the use of computational resources by antivirus programs that are used in the MIS.

2. LITERATURE REVIEW AND PROBLEM POSING

In work [4], that due the digitization of many business processes, many medical institutions have switched to the electronic document management. This was mainly done in the direction of automating the electronic management of medical records [5]. Besides, as noted by many experts [4, 5], a feature of the MIS was the requirement for IS and the DP of patients. In papers [5, 6], it was noted that among cyber threats faced by many ITS in recent years, and in particular MIS, malicious spyware (MS) is a significant proportion [6]. Moreover, in many situations, the architecture of MS is built on the methods of packaging and complication of program code. As noted by the authors [7, 8], the use of classical approaches in problems of detecting of MS is often ineffective.

In a number of publications in recent years, various contexts of problems that arise in the process of recognition of MS were considered. In particular, in papers [8, 9], authors emphasized tasks associated with the formalization of the initial data and the choice of decision procedures in the identification of MS. As it was shown by the authors of [10], in some situations it is quite problematic to draw clear boundaries between algorithms that are used in the case of MS recognition.

In work [11], the existing approaches to the methods of MS recognition were compared. And in the works [8, 9, 11], the incompleteness in study of the question of applying the entropy attributes of a file in the algorithms for detecting MS was emphasized.

Therefore, the task of development of methods, models and algorithms that allows to increase the effectiveness of identifying MS remains relevant in an MIS, which often contain important and sometimes unique information about the condition of patients.

3. GOAL AND OBJECTIVES OF THE STUDY

Improvement of models and algorithms that allow segmentation of binary files. It is proposed to solve this task due to discrete wavelet transforms (DWT).

Within the framework of the article it is necessary:

- to develop a model and algorithm that allow software to implement binary file segmentation using DWT and unlike existing approaches, take into account local extremes for the wavelet coefficients (WC) obtained in the calculations;

- experimentally validate proposed solutions to test the procedure for finding the boundaries of the resulting segments. In this case, the emphasis is on minimizing the requirements for MIS resources.

4. MODELS AND METHODS

The procedure of file segmentation, which is considered in the system of detection of MS, is implemented in two stages. Initially, the file is pre-processed. At the second stage, wavelet analysis with file segmentation is implemented. During the initial processing, the sliding window method was used [7, 8]. This allows us to present the contents of the analyzed file as a series of the following form: $H = \{h_i : i = 1,..., N\}$, where $N -$ number of windows, $h_i -$ information entropy, which is defined for the $i -$ file window [12]. The information entropy for the i - window was calculated using the following formula (taking into account the frequency of fixation of unsimilar bytes in the file) [12]: $h_i = -\sum_{j=1}^{n} p(j)\log_2 p(j)$, where the $p(j) -$ frequency of the occurrence of a $j -$ byte in the $i -$ window, which includes n bytes.

Wavelet analysis was used in the module of identifying MS to divide a series into segments. At the same time, we assume that the segments should have a homogeneous structure. Unlike in [7, 9], it was proposed to apply discrete redundant WT [13]. In general, this stage can be described as:

$$T(a,b) = \frac{1}{\sqrt{a}} \sum_{i=b}^{b+a} h_i \cdot \psi_{HA}\left(\frac{i-b}{a}\right), \quad (1)$$

where the $a, b -$ parameters of scale and shift, respectively; $\psi_{HA} -$ Haar wavelet [13].

By creating calculations redundancy at the expense of, b - you can solve the problem of invariance in the process of shifting the original data. The subsequent implementation of file segmentation can be illustrated using fig. 1. In fig. 1 shown an example of mapping of local

extremes for WC. The task is to find a local extremum at each step of the transformations. The boundaries of the resulting segments are defined in the process of localization of extremes. As a result of the implementation of this stage of identifying the MS, you can get a sequence of certain segments. In addition, each segment can be characterized by its number of bytes.

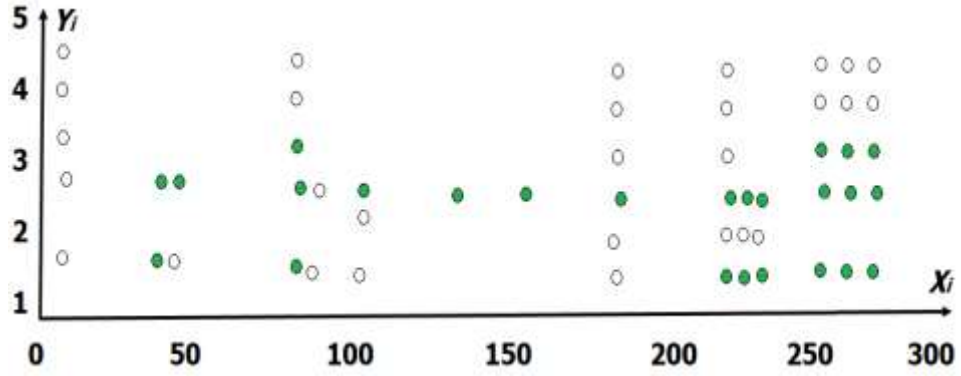

Figure 1. Display of points of local extrema of WC.

Further, we assume that the compared files are described as a sequence of segments. How close files are to each other was determined by calculating the weighted edit distance. To implement the stage and the corresponding file comparison model, the Levenshtein metric was used [13, 14]. However, unlike in the existing models, it was proposed to use the function. θ This function (θ) allowed to determine the amount of "penalty" in a situation when the divergence of two segments was fixed. In this case, the model made the assumption that each of the segments can be characterized using its size and entropy.

$$\Theta = \mu \cdot \frac{|nub_1 - nub_2|}{nub_1 + nub_2} + R, \qquad (2)$$

where, nub_1, nub_2 - respectively, the number of bytes in adjacent segments;

$R = (1-\mu)\cdot(1/(1+\exp(\lambda \cdot |h_1 - h_2| + \omega)))$, λ, μ, ω - empirically selected coefficients for different types of MS.

Given the expression (2), the comparison of two MS's for a function θ was performed using an algorithm that allowed determining values Θ in the procedures for matching sequences of segments. Two MS files are compared. Further, similar segments are found in those files. Each step of the algorithm that allows us to fill the matrix $M[i][j]$, implements the basic operation of the transformation. Under the basic operation for the analyzed file, were meant the following operations: replace, deletion or insertion of a segment. Further it is accepted that: $1 \le i \le |A|$ and $1 \le i \le |B|$ - designations for indices of segments of MS; *A,B* - respectively, the compared sequences.

Step 1. $M[0][0] = 0$,

Step 2. $M[i][0] = M[i-1][0] + \mu \cdot \lg(nub_{A,i-1})$,

Step 3. $M[i][0] = M[i-1][0] + \mu \cdot \lg(nub_{B,i-1})$,,

Step 4. $$M[i+1][j+1] = \begin{cases} M[i][j] + \Theta(A_i, B_i) \cdot \lg((nub_{A,i} + nub_{B,i})/2) \\ M[i][j+1] + \mu \cdot \lg(nub_{A,i}) \\ M[i+1][j] + \mu \cdot \lg(nub_{B,j}) \end{cases},$$

Step 5. Using the methodology described in [15–17], a fuzzy distribution of realizations of objects that are used during training (multi-dimensional sign matrix for training) is performed. The resulting matrices $M[i][j]$ are transformed into a clear distribution when optimizing the test tolerances for each class of MS. As a result, a purposeful change in the values of recognition features for each class of MS occurs to build a correct decision rules in the procedures for identifying specific malware.

Using the methods and algorithms described in [15–17], the algorithm for correcting objects that are used for training the module of identifying MS and training phase itself was also implemented.

During the last phase, the decision rules for the identification of MS are synthesized.

5. SIMULATION EXPERIMENT

To confirm the performance and adequacy of the proposed model and algorithm, experimental studies were conducted. Initially, the estimation of the number of operations required for the implementation of the proposed model and algorithm was performed. At the second phase – the experimental test, a system was tested for the detection of packaged of MS.

The method conducting the experiment involved a test for samples of 3 categories of packaged MS. For the initial tests, the following categories were selected: 1) spyware (own authoring, including Keylogger, which disguises itself as a system process and allows sending information received from an infected computer in the MIS to the malicious user's mail, for example, screenshots, open MIS windows, typed data, correspondence with patients, etc., the code fragment is shown in Fig. 2); 2) MS type Backdoor.Tdss; 3) MS type Trojan.Mayachok. The total number of checks performed - 240 experiments on virtual machines.

On the Figure 3, the results of testing the system for identifying MS are shown.

As can be seen from the results of the analysis of test samples, the developed model and algorithm for identifying MS, quite successfully allowed us to implement procedures for recognizing individual categories of MS. The comparison with a number of common antiviruses that run the top 10 was performed. In particular, the comparison of test results was down with AVG Anti-Virus Free, Avast Antivirus, Panda Antivirus Pro, 360 Total Security, ESET NOD32 Smart Security. These antiviruses are mainly built on signature-based detection approach.

Figure 2. Fragment of the source code of the MS, which was used in the process of testing the module to identify threats.

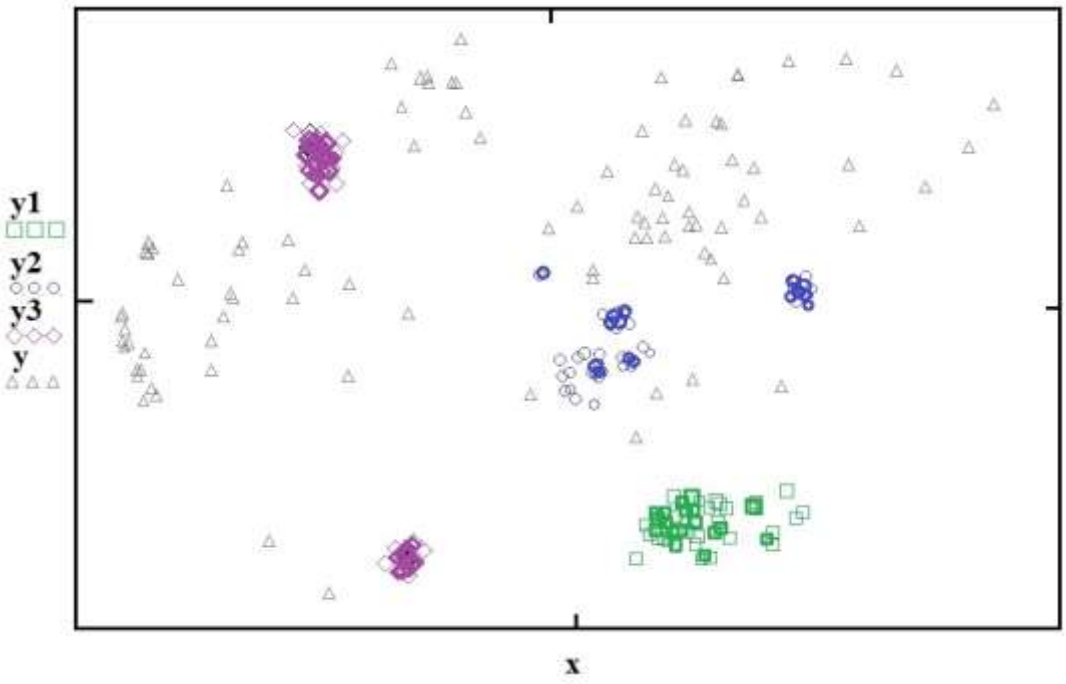

Figure 3. Results of experimental verification of the module for identifying MS.

During the test checks, it was established

The proposed model, the corresponding algorithm and the system for identifying MS showed satisfactory characteristics in recognizing MS during analysis in most experiments (218 out of 240).

The best resistance to obfuscation protection methods was shown in comparison with byte antivirus products. At the same time, the number of errors of the 2nd kind was about 15–18% for the developed system of detection of MS. In comparison with considered antivirus software, this is a fairly good indicator. For example, a spyware program (Windows Security Health Service), which we developed specifically for testing, was detected only by 360 Total Security antivirus.

It was found that the proposed algorithm is less demanding on the number of anti-virus entries. During the test checks, it was found that, for example, only about 50–60 signature records are needed to identify the MS that can be classified as BackDoor.Maxplus. At the same time, the total size of these records was less than 0.2 kBt (which corresponds to three entropy records) in comparison with 0.9–0.98 kBt for the considered antiviruses. Therefore, by reducing the number of anti-virus entries, you can reduce the size of the database of anti-virus programs by 40–45%. This reduces the time for checking potentially dangerous software.

6. RESULTS AND DISCUSSION

The advantages of our study include the following:

1) the proposed solutions, in particular, the additions of models and algorithms that allow software to implement binary file segmentation using DWT, but unlike the results of other authors [6–11] took into account local extremes WC;

2) the model described in the work allows one to take into account not only the number of bytes and the entropy of the segment of MS, but also such a parameter as the value of "recovery" in a situation when there is a discrepancy between the two segments; 3) the proposed solutions were tested experimentally.

In the course of the experiments, the effectiveness of the proposed additions to the models and algorithms for identifying MS was confirmed.

At the present stage of research, a certain lack of work is a small degree of approbation of the proposed solutions. Experiments have so far been carried out only on the platforms of several information systems, including medical ones, which were equipped with the proposed module for the identification of MS.

The prospect of further research is determined by the possibilities of applying the obtained results for the subsequent larger-scale experiment and by testing the module for identifying MS on a larger number of test samples of MS. It is also possible to programmatically automate the processing of data about possible facts of MIS infection with packed MS. In this context, our work continues previous publications [17–27].

7. GRATITUDES

The research and the article were done within the framework of promising scientific and technical programs of the Department of Computer Systems and Networks of the National University of Life and Environmental Sciences of Ukraine, as well as the grant of the Republic of Kazakhstan, registration number AP05132723 “Development of adaptive expert systems in the area of cybersecurity of critical objects of informatization”.

8. CONCLUSIONS

The following results have been obtained:

The additions of the model and algorithm, which allow one to programmatically implement the segmentation of binary files using discrete wavelet transform were proposed. At the same time, unlike the existing approaches, the new model and algorithm took into account local extremes of the wavelet coefficients. The solutions proposed in this paper allow us to take into account in the process of analysis the number of bytes and the entropy of the file segment, which is considered as MS. As an alternative, a solution was considered based on the procedure for analyzing the frequency of occurrence of file bytes in segments. The location of the bytes in the segments was also analyzed. The wavelet transform was chosen on the basis of maximizing efficiency in the analysis of spyware programs that are potentially dangerous for MIS;

In the course of computational experiments, the proposed solutions were tested. The effectiveness of the procedures for the formation of the boundaries of file segments was experimentally confirmed. In the course of experimental studies, emphasis was placed on minimizing the computing resources of information systems, and in particular, medical ones.

REFERENCES

[1] Jiang, Y., Song, H., Wang, R., Gu, M., Sun, J., & Sha, L. (2017). Data-centered runtime verification of wireless medical cyber-physical system. IEEE transactions on industrial informatics, 13(4), pp. 1900–1909.

[2] Sandberg, H., Amin, S., & Johansson, K. H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. IEEE Control Systems, 35(1), pp. 20–23.

[3] Taylor, R. W., Fritsch, E. J., & Liederbach, J. (2014). Digital crime and digital terrorism. Prentice Hall Press.

[4] Jiang, Q., Chen, Z., Li, B., Shen, J., Yang, L., & Ma, J. (2018). Security analysis and improvement of bio-hashing based three-factor authentication scheme for telecare medical

information systems. Journal of Ambient Intelligence and Humanized Computing, 9(4), pp. 1061–1073.

[5] Wager, K. A., Lee, F. W., & Glaser, J. P. (2017). Health care information systems: a practical approach for health care management. John Wiley & Sons.

[6] Lyda, R., & Hamrock, J. (2007). Using entropy analysis to find encrypted and packed malware. IEEE Security & Privacy, 5(2). pp. 40–45.

[7] Tian, R., Islam, R., Batten, L., & Versteeg, S. (2010). Differentiating malware from cleanware using behavioural analysis. In Malicious and Unwanted Software (MALWARE), 2010 5th International Conference on (pp. 23–30). IEEE.

[8] Han, K. S., Lim, J. H., Kang, B., & Im, E. G. (2015). Malware analysis using visualized images and entropy graphs. International Journal of Information Security, 14(1), pp. 1–14.

[9] Bat-Erdene, M., Park, H., Li, H., Lee, H., & Choi, M. S. (2017). Entropy analysis to classify unknown packing algorithms for malware detection. International Journal of Information Security, 16(3), pp. 227–248.

[10] Canfora, G., Mercaldo, F., & Visaggio, C. A. (2016). Anhmm and structural entropy based detector for android malware: An empirical study. Computers & Security, 61, pp. 1–18.

[11] Damodaran, A., Di Troia, F., Visaggio, C. A., Austin, T. H., & Stamp, M. (2017). A comparison of static, dynamic, and hybrid analysis for malware detection. Journal of Computer Virology and Hacking Techniques, 13(1), pp. 1–12.

[12] Borda, Monica (2011). Fundamentals in Information Theory and Coding. Springer. ISBN 978-3-642-20346-6.

[13] Baysa, D., Low, R. M., & Stamp, M. (2013). Structural entropy and metamorphic malware. Journal of computer virology and hacking techniques, 9(4), pp. 179–192.

[14] Apel, M., Bockermann, C., & Meier, M. (2009). Measuring similarity of malware behavior. In Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on (pp. 891–898). IEEE.

[15] Lakhno, V.A., Kravchuk, P.U., Malyukov, V.P., Domrachev, V.N., Myrutenko, L.V., Piven, O.S. (2017). Developing of the cyber security system based on clustering and formation of control deviation signs, Journal of Theoretical and Applied Information Technology, Vol. 95, Iss. 21, pp. 5778–5786.

[16] Lakhno, V., Zaitsev, S., Tkach, Y., Petrenko, T. (2019). Adaptive expert systems development for cyber attacks recognition in information educational systems on the basis of signs' clustering, Advances in Intelligent Systems and Computing, 1st International Conference on Computer Science, Engineering and Education Applications, ICCSEEA2018; Kiev; Ukraine; 18 January 2018, Vol. 754, pp. 673–682

[17] Akhmetov, B., Lakhno, V., Akhmetov, B., Alimseitova, Z. (2019). Development of sectoral intellectualized expert systems and decision making support systems in cybersecurity, Advances in Intelligent Systems and Computing, 2nd Computational Methods in Systems and Software, CoMeSySo 2018; Szczecin; Poland; 12 September 2018, Vol. 860, pp. 162–171.

[18] Lakhno, V., Boiko, Y., Mishchenko, A., Kozlovskii, V., Pupchenko, O. (2017) Development of the intelligent decisionmaking support system to manage cyber protection at the object of informatization. Eastern European Journal of Enterprise Technologies, 9 (86), pp. 53–61.

[19] Tolubko, V., Kozelkov, S., Zybin, S., Kozlovskyi, V., Boiko, Y. (2019). Criteria for evaluating the effectiveness of the decision support system. Advances in Intelligent Systems and Computing,754, pp. 320–330.

[20] Yudin, O., Boiko, Y., Frolov, O. (2015). Organization of decision support systems for crisis management. Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), 2015, pp. 115–117.

[21] Akhmetov, B., etc. (2018). Models and Algorithms of Vector Optimization in Selecting Security Measures for Higher Education Institution's Information Learning Environment. In Proceedings of the Computational Methods in Systems and Software (pp. 135–142). Springer, Cham.

[22] Lakhno, V., Petrov, A., & Petrov, A. "Development of a Support System for Managing the Cyber Security of Information and Communication Environment of Transport", In International Conference on Information Systems Architecture and Technology, 2017, pp. 113–127. Springer, Cham.

[23] Lakhno, V.A., Tretynyk, V.V. (2019) Information Technologies for Maintaining of Management Activity of Universities. In: Hu Z., Petoukhov S., Dychka, I., He, M. (eds) Advances in Computer Science for Engineering and Education. ICCSEEA 2018. Advances in Intelligent Systems and Computing, vol 754. pp. 663–672.

[24] Lakhno, V., Tkach, Y., Petrenko, T., Zaitsev, S. & Bazylevych, V. (2016). Development of adaptive expert system of information security using a procedure of clustering the attributes of anomalies and cyber attacks. Eastern European Journal of Enterprise Technologies, 6/9 (84), pp. 32–44.

[25] Lakhno, V., Akhmetov, B., Korchenko, A., Alimseitova, Z., Grebenuk, V. (2018). Development of a decision support system based on expert evaluation for the situation center of transport cybersecurity. Journal of Theoretical and Applied Information Technology, 96 (14), pp. 4530–4540.

[26] Lakhno, V., Buriachok, V., Parkhuts, L. etc. (2018). Development of a conceptual model of adaptive access rights management with using the apparatus of Petri nets. International Journal of Civil Engineering & Technology (IJCIET), Vol. 9, Iss. 11, pp. 95–104.

[27] Akhmetov, B., Kydyralina, L. etc. (2018). Model for a computer decision support system on mutual investment in the cybersecurity of educational institutions, International Journal of Mechanical Engineering & Technology (IJMET), Vol. 9, Iss. 10, pp. 1114–1122.